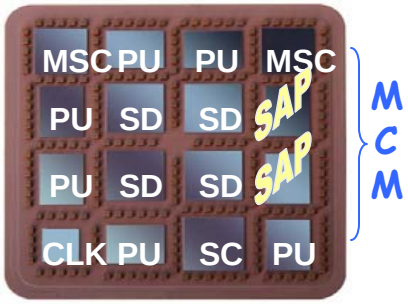


Model	Books	SAPs
S08	1	2
S18	2	4
S28	3	6
S38	4	8
S54	4	8

model defaults

The max number of additional SAPs is 2 per book

ZTIDBIT #17

Cheat Sheet

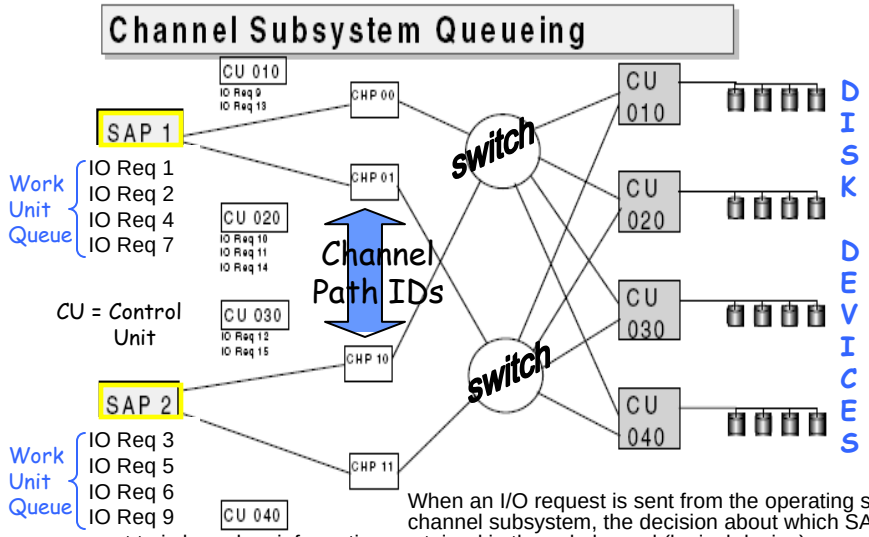
When the MVS/Extended Architecture (MVS/XA) was made available in 1983, the responsibility for selecting channel paths for driving I/O requests was moved from the input/output supervisor (IOS) portion of the operating system to a new component of the hardware known as the channel subsystem (CSS). This offloaded the work that had previously been done by the central processing unit (CPU) to new processing units in the channel subsystem known as I/O processors (IOPs). This design came about because many input/output tasks were fairly complex and require logic to be applied to the data to convert frame formats suitable to do an I/O.
In these situations, the computer's CPU would normally be asked to handle the logic, but due to the fact that the I/O devices are slow, the CPU would end up spending a huge amount of time sitting idle waiting for the data from the device. System z' channel architecture avoids this problem by using a no-cost, but very effective processor with enough logic and memory onboard to handle the I/O role (form of a co-processor). The mainframe calls this a **System Assist Processor** (SAP). This I/O engine is used by all operating systems offered on the mainframe.

A SAP runs special I/O Licensed Internal Code (LIC). The SAP takes responsibility for the processing during the execution of an I/O operation, freeing up the operating system CPs to do other work. It schedules an I/O operation, checking for the full availability of an I/O path, and provides a queue mechanism if an I/O path is not available. The SAP can also prioritizes an I/O in its work queue using an implementation of WorkLoad Management. Therefore, the SAP extends Service Level Objectives down to the device level. Past metrics clocked two billion I/O operations per second for a single channel subsystem. Today z9 can be configured with up to four Channel Subsystems.

Other machine designs use one of their standard general purpose processors to perform the I/O role. This design slows down the entire workload and overall application throughput. This disrupts the kernel or the process supervisor to do the I/O engine selection, schedule the I/O and await its completion to resume the task's work. This incurs overhead (context switching) where work areas, registers and instruction sets are moved around just to do an I/O. Some machine designs have a dedicated logical partition just to do I/O. This is expensive use of resources where the I/O has to take an additional hop to schedule it. Therefore, when a customer purchases a non-z machine, they'll need to factor in I/O workload requirements and purchase additional engines (utilization may likely be unknown), where z provides SAPs for free. They are part of the z-Architecture.

Depending on the processor model (S08, S18….S54), there will be two or more SAPs used in a channel subsystem. You can add more SAPs over its model defaults as the I/O requirements demand. The SAP finds the subchannel (device) in the initiative queue and tries to find a channel that succeeds in *initial selection* (connects to a control unit and starts the I/O operation). The SAP uses information in the subchannel to determine which channels and control units can be used to reach the target device. All this information on channels, control units and device configuration is contained within non-addressable storage called the Hardware Save Area (HSA). – See middle diagram above.

On the current technology, there is an affinity between a channel and a given SAP. Most devices will be connected to multiple channels, and the IBM configuration guidelines recommend configuring devices on channels that are spread across more than one SAP.



When an I/O request is sent from the operating system to the channel subsystem, the decision about which SAP to route the request to is based on information contained in the subchannel (logical device).

Each SAP has a work-unit queue that contains all the Start Subchannel (SSCH) I/O requests from all the LPARs for a channel that is attached to that SAP. The requests in this queue are processed based upon the I/O Priority assigned either by WLM or by the hardware (if the LP does not exploit Channel Subsystem I/O Priority Queueing – feature of IRD).
if there is an available channel to the device on a SAP, the request will be initiated. If there are no available channels on that SAP, the request is re-queued to another SAP which also has connectivity to the target device. The request will then be placed in the work queue for the alternative SAP, and once again is processed based upon its I/O Priority.

Using the Hardware Management Console (HMC) or the machine's Support Element (SE) additional SAPs can be "characterized" . An important point - You can reassign your system SAP allocations to optimize performance of applications without performing a power-on reset of the CPC.
Changing the CPs to SAPs requires reassignment of the current licensed internal code.
You must ensure the the CPs being converted to SAPs can be deconfigured and are not dedicated.

<u>Side Note</u>: SAPs are also used for CPU sparing when transitioning between a failed and new CPU. SAP is used as a temporary hold area for the instruction – registers while new CPU is brought on line.